

On the Rigidity of Sparse Random Graphs

Nati Linial *

Jonathan Mosheiff †

Abstract

A graph with a trivial automorphism group is said to be *rigid*. Wright proved [11] that for $\frac{\log n}{n} + \omega(\frac{1}{n}) \leq p \leq \frac{1}{2}$ a random graph $G \in G(n, p)$ is rigid whp. It is not hard to see that this lower bound is sharp and for $p < \frac{(1-\epsilon)\log n}{n}$ with positive probability $\text{aut}(G)$ is nontrivial. We show that in the sparser case $\omega(\frac{1}{n}) \leq p \leq \frac{\log n}{n} + \omega(\frac{1}{n})$, it holds whp that G 's 2-core is rigid. We conclude that for all p , a graph in $G(n, p)$ is reconstructible whp. In addition this yields for $\omega(\frac{1}{n}) \leq p \leq \frac{1}{2}$ a canonical labeling algorithm that almost surely runs in polynomial time with $o(1)$ error rate. This extends the range for which such an algorithm is currently known [5].

1 Introduction

It is a truth universally acknowledged, that random objects are asymmetric. It was shown by Wright [11] that for $\frac{1}{2} \geq p \geq \frac{\log n}{n} + \omega(\frac{1}{n})$ a random $G(n, p)$ graph has, whp, a trivial automorphism group. He actually worked with the $G(n, M)$ model, but the reduction to $G(n, p)$ is well-known and follows easily from the Chernoff bound. Also, a graph and its complement clearly have the same automorphism group, so we can restrict ourselves to the range $\frac{1}{2} \geq p$. Wright's bound is tight, since a graph G of slightly smaller density is likely to have isolated vertices, which can be swapped by a G -automorphism. This paper concerns the range of smaller p by showing that for $\omega(\frac{1}{n}) \leq p \leq n^{-\frac{1}{2}-\epsilon}$ whp all of G 's automorphisms are essentially trivial. Here is our main result:

Theorem 1. *Let $G = (V, E)$ be a $G(n, p)$ graph with $\omega(\frac{1}{n}) \leq p \leq n^{-\frac{1}{2}-\epsilon}$. Then whp its 2-core has a trivial automorphism group.*

*Department of Computer Science, Hebrew University, Jerusalem 9190401. e-mail: nati@cs.huji.ac.il. Supported by a grant from the Israel Science Foundation.

†Department of Computer Science, Hebrew University, Jerusalem 9190401. e-mail: yonatanm@cs.huji.ac.il. Supported by the Adams Fellowship Program of the Israel Academy of Sciences and Humanities.

This shows that for this range of p , $\text{whp aut}(G)$ is generated by:

- Automorphisms of rooted trees that are attached to the 2-core.
- Automorphisms of the tree components and swaps of such components.

The most interesting range of this statement is $p \leq \frac{\log n + (1+\epsilon) \log \log n}{n}$. For larger p the 2-core is the whole graph, in which range ours is just a new proof for the rigidity of sufficiently dense random graphs.

General strategy of the proof: We denote the vertex set of G 's 2-core by $R(G)$. It is easy to see that $\text{aut}(G)$ fixes $R(G)$ setwise and our proof shows first that $\text{aut}(G)$ actually fixes $R(G)$ pointwise. In order to prove the theorem in full we show that this rigidity does not result from boundary effects of vertices near $V \setminus R(G)$. The neighbor set of $v \in V$ and its degree are denoted by $N(v)$ and $d(v)$. If x_1, \dots, x_k are the neighbors of v , we denote by $\nabla(v)$ the *multiset* $\{d(x_i)\}_1^k$. Clearly ∇ is preserved by automorphisms. We fix some $k \leq \log n$ and consider two directed rooted cycles v_1, \dots, v_k and u_1, \dots, u_k in G . We show that whp every two such cycles have many *incompatible* pairs (v_i, u_i) for which $\nabla(v_i) \neq \nabla(u_i)$. This already implies that $R(G)$ is fixed pointwise. In the full proof of the theorem we find, for every two such cycles, an incompatible pair (v_i, u_i) , where both v_i and u_i are at distance ≥ 3 from $V \setminus R(G)$. Such a pair is not only incompatible in G , but also in $R(G)$, proving the theorem.

It turns out that Theorem 1 yields some interesting insights on the well-known graph reconstruction conjecture which we now recall. Let G be an n -vertex graph. When we delete a vertex of G we obtain an $(n - 1)$ -vertex graph. By doing this separately for each vertex in G we obtain the n graphs that make up G 's *deck*. The *graph reconstruction conjecture* ([7], [10]) posits that every two graphs of 3 or more vertices that have identical decks must be isomorphic. A graph G is said to be *reconstructible* if every graph with the same deck is isomorphic to G . Bollobás proved [2] that $\text{whp } G(n, p)$ graphs are reconstructible for all $\frac{(5/2+\epsilon) \log n}{n} \leq p \leq 1 - \frac{(5/2+\epsilon) \log n}{n}$. We show that this is in fact true for every $0 \leq p \leq 1$. One reason why this extension of range is of interest has to do with the *edge reconstruction conjecture* [6] which states that every graph can be reconstructed from its deck of edge-deleted subgraphs. This leads to the notion of *edge-reconstructible* graphs. We recall two facts from this theory: (i) Every reconstructible graph with no isolated vertices is edge-reconstructible (e.g., [3]) (ii) Every n -vertex graph with at least $\log_2(n!) + 1 = n \log_2 n + O(n)$ edges is edge reconstructible. Our result applies to the range $|E| \leq O(n \log n)$ where the edge reconstruction problem is still open.

We turn to discuss the *canonical labeling problem* [1]. Let \mathcal{L} be a class of graphs. A canonical labeling of $G \in \mathcal{L}$ assigns distinct labels to the vertices of G , where the labeling is uniquely determined by G 's isomorphism class. In the

probabilistic version of this problem, \mathcal{L} is a probability space of graphs and we seek to efficiently find a canonical labeling for almost all graphs in \mathcal{L} . Such a canonical labeling algorithm clearly solves in particular the *random graph isomorphism problem* for \mathcal{L} . Specifically we ask for which values of p there is a polynomial time canonical labeling in $G(n, p)$. By considering the complementary graph it suffices to consider the range $p \leq 1/2$. Such an algorithm is known [5] for $p \in [\Theta(\frac{\ln n}{n}), 1/2]$. Our proof of Theorem 1 yields a polynomial time algorithm for $\omega(\frac{1}{n}) \leq p \leq n^{-(0.5+\epsilon)}$, whence a polynomial time solution exists for $p \in [\omega(\frac{1}{n}), \frac{1}{2}]$.

2 Technical Preliminaries

Graph theory: Graphs are denoted $G = (V, E)$ and usually $n := |V|$. The neighbor set of $u \in V$ is denoted by $N(u)$. For $U \subseteq V$, we denote $N(U) := (\bigcup_{u \in U} N(u)) \setminus U$ and $\tilde{N}(U) := \bigcup_{u \in U} \tilde{N}(u)$.

The set of cross edges between two subsets $U, W \subseteq V$ is denoted $E(U, W) := \{uv \in E \mid u \in U, v \in W\}$, and $d(U, W) = |E(U, W)|$ (to wit: even if $U \cap W \neq \emptyset$, we consider every relevant edge exactly once). For a singleton $U = \{u\}$, we use the shorthand $d(u, W) = d(U, W)$. Also, $E(U) = E(U, U)$.

For $U \subseteq V$ we denote $\sigma(U) := \{v \in V \setminus U \mid d(v, U) = 1\}$, the set of those vertices not in U that have exactly one neighbor in U .

We denote by G_U the subgraph of G induced by $U \subseteq V$.

Let $\nabla(u)$ denote the *multiset of integers* $\{d(v, V \setminus \tilde{N}(u)) \mid v \in N(u)\}$.

We denote the vertex set of G 's 2-core by $R(G)$.

Asymptotics: A property of $G(n, p)$ graphs is said to hold *whp* (with high probability) if its probability tends to 1 as $n \rightarrow \infty$.

Probability: For a discrete random variable X , let

$$\Pi(X) = \sup_{x \in \text{range}(X)} (\Pr(X = x)).$$

If X is multinomial with parameters $(m, (p_1, \dots, p_k))$, we denote $\Pi(X)$ by $\Pi(m, (p_1, \dots, p_k))$. The following lemmas provide a description of $\Pi(m, (p_1, \dots, p_k))$.

Lemma 2. *Let X be a multinomial random variable with parameters $(m, (p_1, \dots, p_k))$ and suppose that $\Pi(X) = \Pr(X = (a_1, \dots, a_k))$. Then, $a_t > mp_t - 1$ for every t , or, in other words $a_t \geq \lfloor mp_t \rfloor$.*

Proof. Without loss of generality, assume by contradiction that $a_1 \leq m \cdot p_1 - 1$. Then, since $\sum_i a_i = m = \sum_i m \cdot p_i$, there exists some index s , say $s = 2$ such that $a_s > m \cdot p_s$.

$$\frac{\Pr(X = (a_1 + 1, a_2 - 1, a_3, \dots, a_k))}{\Pr(X = (a_1, \dots, a_k))} = \frac{a_2}{a_1 + 1} \cdot \frac{p_1}{p_2} > \frac{m \cdot p_2}{m \cdot p_1} \cdot \frac{p_1}{p_2} = 1$$

contrary to the assumed maximality of $\Pr(X = (a_1, \dots, a_k))$. \square

Lemma 3. For an integer m , a constant $c > 0$ and a probability vector $\mathbf{p} = (p_1, \dots, p_k)$, such that $p_i \leq \frac{c}{\sqrt{m}}$ for each i , it holds that

$$\Pi(m, \mathbf{p}) \leq m^{-\Omega\left(\frac{\sqrt{m}}{c}\right)}.$$

Proof. We first show how to reduce the proof to the case where $p_i \geq \frac{c}{3\sqrt{m}}$ for each i . Assume that the lemma holds in this case. For a real vector \mathbf{u} and two coordinate indices $i \neq j$, let $\mathbf{u}_{i,j}$ be the vector obtained by eliminating the coordinates u_i, u_j and introducing a new coordinate of $u_i + u_j$. Let X and $X_{i,j}$ be multinomial random variables with parameters (m, \mathbf{p}) , $(m, \mathbf{p}_{i,j})$, respectively. Note that for every $\mathbf{a} \in \text{range}(X)$ there holds

$$\Pr(X = \mathbf{a}) \leq \Pr(X_{i,j} = \mathbf{a}_{i,j}).$$

Thus, $\Pi(m, \mathbf{p}) \leq \Pi(\mathbf{p}_{i,j})$.

We generate a sequence of probability vectors that start from \mathbf{p} and proceed as follows. At each step we replace, as described, the two smallest coordinates in the present probability vector by one coordinate that is their sum. We continue with this process until the first time at which this vector \mathbf{q} has at most one coordinate that is smaller than $\frac{c}{2\sqrt{m}}$. If the smallest coordinate in \mathbf{q} is $\geq \frac{c}{3\sqrt{m}}$, then, since each of the above steps can only increase Π , the reduction is complete. Otherwise, \mathbf{q} has exactly one coordinate, say q_1 that is $< \frac{c}{2\sqrt{m}}$. But then in $\mathbf{q}_{1,2}$ all coordinates vary between $\frac{c}{2\sqrt{m}}$ and $\frac{3c}{2\sqrt{m}}$. The reduction is again complete.

We now turn to proving the lemma for the case where $\frac{c}{\sqrt{m}} \geq p_i \geq \frac{c}{3\sqrt{m}}$ for $i = 1, \dots, k$. Clearly $k \geq \frac{\sqrt{m}}{c}$. Set $\mu_i = p_i \cdot m$ and suppose that $\Pi(X) = \Pr(X = (a_1, \dots, a_k))$. By Lemma 2, $\frac{a_i}{\mu_i} \geq 1 - \frac{1}{\mu_i} \geq 1 - \frac{3}{c\sqrt{m}}$ for all i . Now

$$\Pi(X) = \Pr(X = (a_1, \dots, a_k)) = \binom{m}{a_1, \dots, a_k} \cdot \prod_i p_i^{a_i}.$$

By Stirling's bound, $\frac{n!}{\left(\frac{n}{e}\right)^n \sqrt{2\pi n}} = 1 + O\left(\frac{1}{n}\right)$. Thus,

$$\Pi(X) \leq O\left(\frac{\left(\frac{m}{e}\right)^m \sqrt{2\pi m}}{\prod_i \left(\frac{a_i}{e \cdot p_i}\right)^{a_i} \cdot \sqrt{2\pi a_i}}\right),$$

which can be stated as

$$\Pi(X) \leq O \left(\frac{\left(\frac{m}{e}\right)^m \sqrt{2\pi m}}{\prod_i \left(\frac{\mu_i}{e \cdot p_i}\right)^{a_i} \cdot \sqrt{2\pi a_i}} \cdot \prod_i \left(\frac{\mu_i}{a_i}\right)^{a_i} \right) = O \left(\frac{\sqrt{2\pi m}}{\prod_i \sqrt{2\pi a_i}} \cdot \prod_i \left(\frac{\mu_i}{a_i}\right)^{a_i} \right).$$

But

$$\prod_i \left(\frac{\mu_i}{a_i}\right)^{a_i} \leq \left(1 + \frac{4}{c\sqrt{m}}\right)^m \leq e^{\frac{4\sqrt{m}}{c}}$$

and

$$\frac{1}{\prod_i \sqrt{a_i}} \leq \left(\frac{c}{3}\sqrt{m} - 1\right)^{-\frac{k}{2}} \leq \left(\frac{c}{3}\sqrt{m} - 1\right)^{-\frac{\sqrt{m}}{2c}} \leq O \left(\left(\frac{c^2 m}{9}\right)^{-\frac{\sqrt{m}}{4c}} \right).$$

Therefore,

$$\Pi(X) \leq O \left((2\pi)^{-k/2} \sqrt{m} \cdot e^{\frac{4\sqrt{m}}{c}} \cdot \left(\frac{c^2 m}{9}\right)^{-\frac{4\sqrt{m}}{c}} \right) \leq m^{-\Omega\left(\frac{\sqrt{m}}{c}\right)}.$$

□

Lemma 4. *Let k be an integer, $\frac{1}{2} \geq p > 0$ and let $p_i = \binom{k}{i} p^i q^{k-i}$ ($i = 0, 1, \dots, k$), where $q = 1 - p$. Then, for every $m \leq O(kp)$ there holds*

$$\Pi(m, (p_0, \dots, p_k)) \leq m^{-\Omega(\sqrt{m})}.$$

Proof. It is well known that

$$\Pi(k, (p, q)) \leq O \left(\left(\frac{1}{pk-1} \right)^{1/2} \right) \leq O \left(\frac{1}{\sqrt{m}} \right).$$

Therefore, by Lemma 3,

$$\Pi(m, (p_0, \dots, p_k)) \leq m^{-\Omega(\sqrt{m})}.$$

□

3 The Main Theorem

We recall that $R(G)$ stands for G 's 2-core. We also denote $\tilde{R} := V \setminus R(G)$.

Lemma 5. *Let G be a $G(n, p)$ graph where $p > \omega(\frac{1}{n})$. For every $\frac{n}{10} > x > \frac{n}{e^{np}}$ there holds*

$$\Pr(|\tilde{R}| \geq x) < e^{-\Omega(np x)}.$$

Proof. Let $S \subseteq V$ be the set of those vertices in G with degree at most 3. We claim that $|S| \geq \frac{|\tilde{R}|}{4}$. Clearly $|E(\tilde{R})| < |\tilde{R}|$, since \tilde{R} is acyclic. Also, $d(\tilde{R}, R) \leq |\tilde{R}|$ since a vertex in \tilde{R} can have at most one neighbor in R . Hence,

$$4|\tilde{R}| - 4|S| \leq 4|\tilde{R} \setminus S| \leq \sum_{v \in \tilde{R}} d(v) = 2|E(\tilde{R})| + d(\tilde{R}, R) < 3|\tilde{R}|,$$

as claimed. Thus, it is enough to bound the probability that $|S| \geq \frac{1}{4}x$. We fix a set A of $\frac{x}{4}$ vertices and note that a vertex $v \in A$ has $d(v) < 4$ only if $d(v, V \setminus A) < 4$, which holds with probability $\leq e^{-\Omega(np)}$. Thus, the probability that all vertices in A have degree ≤ 4 is at most $e^{-\Omega(np x)}$. Therefore, the probability that such a set A exists is at most

$$\binom{n}{\frac{1}{4}x} e^{-\Omega(np x)} = e^{-\Omega(np x)}$$

finishing the proof. \square

Definition 6. *Let $G = (V, E)$ be an n -vertex graph, and $k \geq 3$ an integer. An order k configuration of G is a pair of functions $(\phi, \psi) : [k] \rightarrow V$. If $\phi(i) = \psi(i)$ we say that i is a confluence of (ϕ, ψ) .*

- *A confluence-free configuration (ϕ, ψ) is said to have type I when $k \leq \log n$ and $(\phi(1), \dots, \phi(k), \phi(1))$ and $(\psi(1), \dots, \psi(k), \psi(1))$ are simple cycles (in this order).*
- *We say that (ϕ, ψ) is a type II configuration when $(\phi(1), \dots, \phi(k))$ and $(\psi(1), \dots, \psi(k))$ are each a simple path or a simple cycle. Also, $k \leq \log n$, and 1, k are the only confluences.*

Lemma 7. *Let $G = (V, E)$ be a random $G(n, p)$ graph and let $k \leq \log n$. Pick the functions $\phi, \psi : [k] \rightarrow V$ uniformly at random. Consider the events*

- C_1 that (ϕ, ψ) is a type I configuration.
- C_2 that (ϕ, ψ) is a type II configuration.

Then:

1. $\Pr(C_1) \leq p^k \cdot (\frac{2}{n} + p)^k$
2. $\Pr(C_2) \leq p^k \cdot (\frac{2}{n} + p)^{k-2} \cdot n^{-2}$

Proof. We only prove the first claim. The same argument applies as well to the second case.

Denote $\phi(k+1) = \phi(1)$ and $\psi(k+1) = \psi(1)$. For $i = 0, \dots, k$, we estimate the probability of the events A_i that $\psi(j)\psi(j+1) \in E$ for every $1 \leq j \leq i$ and $(\phi(1), \dots, \phi(k+1))$ is a simple cycle in G . Clearly, $\Pr(A_0) = \Pr(\phi(1), \dots, \phi(k+1) \text{ is a simple cycle}) \leq p^k$.

We complete the proof by showing that $\Pr(A_{i+1}|A_i) \leq \frac{2}{n} + p$. Indeed, suppose that $\psi(i) = \phi(j)$ for some j . In this case it is possible that $\psi(i), \psi(i+1)$ are neighbors since $\psi(i+1)$ coincides with either $\phi(j-1)$ or with $\phi(j+1)$, but that happens with probability $\leq \frac{2}{n}$. Otherwise, they are neighbors with probability p . \square

Lemma 8. Let $\omega(\frac{1}{n}) \leq p = p(n) \leq O(n^{-0.5-\epsilon})$ for some $\epsilon > 0$. Pick a random $G(n, p)$ graph $G = (V, E)$ and two random maps $\phi, \psi : [k] \rightarrow V$ where $k \leq \log n$. Let s denote the number of indices $i \in \{1, \dots, k\}$ such that $\nabla(\phi(i)) = \nabla(\psi(i))$. Then:

$$\Pr\left(s > \frac{1}{4}k \mid C_1\right) \leq (pn)^{-\Omega(\sqrt{pn} \cdot k)}$$

$$\Pr\left(s > \frac{1}{4}(k-2) \mid C_2\right) \leq (pn)^{-\Omega(\sqrt{pn} \cdot k)}$$

Proof. We only prove the type I case. The same argument applies to type II configurations as well. The argument below and all relevant calculations take place in the space conditioned on C_1 .

Let $T = \text{Image}(\phi) \cup \text{Image}(\psi)$ and $t = |T|$. For each index i let U_i be the set of those neighbors of $\phi(i)$ that have no other neighbor in $\tilde{N}(T)$. We expose the subgraph induced on $\tilde{N}(T)$, thus revealing the sets U_i . The following proposition comes in handy:

Proposition 9. With probability $1 - e^{-\Omega(np t)}$ there holds:

- $|\tilde{N}(T)| < 2npt$
- There are at least $\frac{7k}{8}$ indices $k \geq i \geq 1$ for which $\frac{np}{4} \leq |U_i| \leq 4np$.

We proceed under the conditioning that the conclusion of this Proposition holds. We next reveal the edges connecting $\tilde{N}(T) \setminus \bigcup_i U_i$ and $V \setminus \tilde{N}(T)$. This determines $\nabla(\psi(j))$ for all j . On the other hand, $\nabla(\phi(i))$ is completely determined by the neighbor sets of vertices from U_i in $V \setminus \tilde{N}(T)$. Consequently the family of multisets $\{\nabla(\phi(i))\}_i$ is independent.

We are concerned with the event that $\nabla(\phi(i)) = \nabla(\psi(i))$. At this stage this may already be impossible, and if possible, this uniquely determines the multiset

of degrees $d(x, V \setminus \tilde{N}(T))$ over $x \in U_i$. The elements of this multiset are drawn from a binomial distribution, so by Lemma 4, if $\frac{np}{4} \leq |U_i| \leq 4np$, then

$$\Pr(\nabla(\phi(i)) = \nabla(\psi(i))) \leq (np)^{-\Omega(\sqrt{np})}.$$

Note that for $s > \frac{1}{4}k$ to hold, the equality $\nabla(\phi(i)) = \nabla(\psi(i))$ must hold for at least $\frac{k}{8}$ of the indices i for which $|U_i| \geq \frac{np}{4}$. Hence,

$$\Pr(s > \frac{1}{4}k) \leq \binom{\frac{7}{8}k}{\frac{1}{8}k} (np)^{-\Omega(k\sqrt{np})} \leq (np)^{-\Omega(k\sqrt{np})},$$

as stated.

Proof of Proposition 9: The first claim follows from Chernoff's bound, as we observe that

$$|\tilde{N}(T)| \sim t + \text{Bin}(n - t, 1 - q^t), \text{ where } q = 1 - p.$$

so that

$$|\mathbb{E}(|\tilde{N}(T)|)| \leq npt(1 + o(1)).$$

For the second claim

$$|\sigma(T)| \sim \text{Bin}(n - t, tpq^{t-1})$$

and so

$$\mathbb{E}(|\sigma(T)|) \geq npt(1 - o(1)).$$

Let A denote the event that $|\tilde{N}(T)| \leq 2npt$ and $|\sigma(T)| \geq \frac{npt}{2}$. By Chernoff's bound,

$$\Pr(A) \geq 1 - e^{-\Omega(npt)}.$$

Now,

$$d(\sigma(T), N(T)) \sim \text{Bin}\left(|\sigma(T)| \cdot \left(|\tilde{N}(T)| - t - \frac{|\sigma(T)| + 1}{2}\right), p\right)$$

Let B denote the event that A holds, and, in addition, $d(\sigma(T), N(T)) < \frac{npt}{8}$. Note that A implies

$$|\sigma(T)| \cdot \left(|\tilde{N}(T)| - t - \frac{|\sigma(T)| + 1}{2}\right) \leq |\tilde{N}(T)|^2 \leq (2npt)^2.$$

Hence,

$$\Pr(B|A) \geq 1 - \binom{(2npt)^2}{\frac{npt}{8}} p^{\frac{npt}{8}} \geq 1 - e^{-\Omega(npt)}$$

and so

$$\Pr(B) = \Pr(B|A) \cdot \Pr(A) \geq 1 - e^{-\Omega(np t)}.$$

Let $U = \bigcup_i U_i$. Note that B implies that at least $\frac{np t}{4}$ vertices in $\sigma(T)$ have no neighbor in $N(T)$, and thus, $|U| \geq \frac{np t}{4}$. Clearly, $|U_i| \sim \text{Bin}(|U|, \frac{1}{t})$. Let D_i denote the event that $|U_i| < \frac{np}{4}$. For $x \geq \frac{np t}{4}$, Chernoff's bound implies

$$\Pr(D_i \mid |U| = x) \leq e^{-\Omega(np)}.$$

Note that given $|U| = x$, the event D_i is negatively correlated with every event of the form $\bigcap_{j \in J} D_j$ where $\emptyset \neq J \subseteq [k] \setminus i$. Thus, for every $I \subseteq [k]$,

$$\Pr\left(\bigcap_{i \in I} D_i \mid |U| = x\right) \leq \prod_{i \in I} \Pr(D_i \mid |U| = x) \leq e^{-\Omega(np|I|)}.$$

In particular, the event \tilde{D} that at most $\frac{k}{16}$ of the D_i hold satisfies

$$\Pr(\tilde{D} \mid |U| = x) \geq 1 - \binom{k}{k/16} \cdot e^{-\Omega(np t)} \geq 1 - e^{-\Omega(np t)}$$

which implies

$$\Pr(B \cap \tilde{D}) \geq 1 - e^{-\Omega(np t)}.$$

Let F be the event that at most $\frac{k}{16}$ of the U_i 's satisfy $|U_i| < 4np t$. A similar argument shows that given

$$\Pr(B \cap \tilde{F}) \geq 1 - e^{-\Omega(np t)}$$

and we conclude that

$$\Pr(B \cap \tilde{D} \cap \tilde{F}) \geq 1 - e^{-\Omega(np t)}.$$

□

At this stage we have already established the following whp: For $\omega(\frac{1}{n}) \leq p(n) \leq O(n^{-0.5-\epsilon})$ every automorphism of a $G(n, p)$ graph pointwise fixes its 2-core. However, we seek to prove the stronger statement that the 2-core has no nontrivial symmetries. As before consider two random maps $\phi, \psi : [k] \rightarrow V$ where $3 \leq k \leq \log n$. Let $T = \text{Image}(\phi) \cup \text{Image}(\psi)$ and define the events C_1, C_2 as above. Clearly $T \subseteq R$, since T is a union of cycles, and now we need to control the effect of non-2-core vertices on $\text{aut}(G)$. This effect is mediated by the set $P \subseteq T$ of T 's *peripheral* vertices, namely those within distance 2 of $\tilde{R} = V \setminus R$. As we show, the above-mentioned effect is not large, since $|P|$ tends to be small. We prove

Lemma 10.

$$\Pr\left(|P| > \frac{k}{8} \mid C_1\right) \leq (np)^{-\omega_n(k)}$$

$$\Pr\left(|P| > \frac{k}{8} \mid C_2\right) \leq (np)^{-\omega_n(k)}$$

Proof. We only prove case I. The same argument applies as well to case II. All our arguments below are made *conditioned on* C_1 . Let $q = 1 - p$ and $t = |T|$. Clearly, $k \leq t \leq 2k$.

Reveal the subgraph H of G , induced by $V \setminus T$. Denote $W = V(H) \setminus R(H)$. Let $x = \frac{n}{(np)^{\log(np)}}$. By Lemma 5,

$$\Pr(|W| \leq x) \geq 1 - e^{-\Omega(np x)} \geq 1 - (np)^{-\omega(k)}.$$

We henceforth condition on this event. Note that $\tilde{R} \subseteq W$, and thus, it is enough to bound the number of vertices in T at distance ≤ 2 from W . We denote $N_H(W)$ by Q . We claim that $|Q| \leq |W|$, since every vertex in Q has a neighbor in W , whereas every vertex in W has at most one neighbor in Q . (Note that $Q \subseteq R(H)$ and a vertex with more than one neighbor in Q is in $R(H)$ as well).

To understand the set P of peripheral vertices, we define three sets P_1, P_2, P_3 with $P \subseteq P_1 \cup P_2 \cup P_3$ and show that whp all $|P_i|$ are small. Let P_1 be the set of those vertices in T with a neighbor in W . Let P_2 be the set of those vertices in T with a neighbor in P_1 . Finally, P_3 is the set of those vertices in T with a neighbor in Q .

Now reveal the set of cross edges $E(T, V \setminus T)$. For $v \in T$, the probability that v has a neighbor in W is at most xp . Thus,

$$\Pr\left(|P_1| \geq \frac{k}{400}\right) \leq \binom{t}{\frac{k}{400}} (xp)^{\frac{k}{400}} \leq (np)^{-\Omega(k \log(np))}$$

and similarly, $\Pr(|P_3| \geq \frac{k}{400}) \leq (np)^{-\Omega(k \log(np))}$. In what follows we condition on the event that $|P_1|, |P_3| \leq \frac{k}{400}$.

We finish by bounding $|P_2|$. Reveal the edge set $E(P_1, T)$. By assumption, T is the image of a type I configuration, namely two simple cycles, possibly with some overlaps. This implies the existence of certain edges in $E(P_1, T)$, at most $4|P_1|$ in number. In addition, the random variable $d(P_1, T)$ is a sum of at most $|P_1||T|$ independent Bernoulli- p random variables. By assumption $4|P_1| \leq \frac{k}{100}$, so that $d(P_1, T) > \frac{k}{50}$ only if at least $\frac{k}{100}$ of these Bernoulli trials succeed. Therefore

$$\Pr\left(d(P_1, T) > \frac{k}{50}\right) \leq \binom{\frac{kt}{400}}{\frac{k}{100}} \cdot p^{\frac{k}{100}} \leq (kp)^{\Omega(k)} \leq (np)^{-\omega_n(k)}$$

Clearly, $|P_2| \leq d(P_1, T)$, and so, $|P_2| \leq \frac{k}{50}$ with probability at least $1 - (np)^{-\omega_n(k)}$. \square

Definition 11. A configuration (ϕ, ψ) of G is said to be compatible if there exists an automorphism π of $R(G)$ such that $\pi(\phi(i)) = \psi(i)$ for each i .

Lemma 12. Let G be a random $G(n, p)$ graph with $\omega(\frac{1}{n}) \leq p \leq O(\frac{\log n}{n})$. Then whp G contains no compatible configuration of type I or II.

Proof. We prove the claim for type I configurations. The proof for type II follows the same argument.

In the coming paragraph we denote $\nabla(v)$ by $Y(v)$. We also consider the 2-neighborhood of $v \in R(G)$ in the subgraph induced by $R(G)$ and denote $\nabla(v)$ in that graph by $Z(v)$. Clearly, a configuration $\phi, \psi : [k] \rightarrow V$ can be compatible only if $Z(\phi(i)) = Z(\psi(i))$ for each $1 \leq i \leq k$.

Let $3 \leq k \leq \log n$ and pick two functions $\phi, \psi : [k] \rightarrow V$ uniformly at random. By Lemma 7, the probability that (ϕ, ψ) is a configuration is at most $p^{k+o_n(1)}$. Conditioned on this event, let

$$A = \{i \in [k] \mid Y(\phi(i)) \neq Y(\psi(i))\}.$$

By Lemma 8, $\Pr(|A| < \frac{3}{4}k) \leq (np)^{-\omega_n(k)}$. Let

$$B = \{i \in [k] \mid Y(\phi(i)) = Z(\phi(i)) \text{ and } Y(\psi(i)) = Z(\psi(i))\}.$$

Note that $i \in B$ when both $\phi(i)$ and $\psi(i)$ are non-peripheral. Hence, by Lemma 10, $\Pr(|B| < \frac{3}{4}k) \leq (np)^{-\omega_n(k)}$. But $|A|, |B| \geq \frac{3}{4}k$, so they must intersect, say $i \in A \cap B$. Then $Z(\phi(i)) \neq Z(\psi(i))$, which makes (ϕ, ψ) incompatible. Clearly this holds with probability $1 - (np)^{-\omega_n(k)}$.

If a_k is the number of compatible type I configurations we can now estimate its expectation:

$$\mathbb{E}(a_k) = n^{2k} \cdot p^{2k+o_n(1)} \cdot (np)^{-\omega_n(k)} \leq (np)^{-\omega_n(k)}$$

and so

$$\sum_{k=3}^{\log n} \mathbb{E}(a_k) \leq (np)^{-\omega_n(1)}$$

which completes the proof. \square

We can now finish up the proof of our main theorem.

Theorem 1. Let $G = (V, E)$ be a $G(n, p)$ graph with $\omega(\frac{1}{n}) \leq p \leq n^{-\frac{1}{2}-\epsilon}$. Then whp its 2-core has a trivial automorphism group.

Proof. Let H denote the 2-core of G . It is known ([4]) that whp $\text{diam}(G) < \frac{\log n}{2}$, which we henceforth assume.

Suppose that $\pi(v) \neq v$ for some $\pi \in \text{aut}(H)$, and a vertex $v \in R(G)$. By Lemma 12, it is enough to show that this assumption implies that G has a compatible configuration. It is easy to see that if π fixes all vertices of H contained in cycles, then π is trivial, so let C be a cycle that contains v . The bound on G 's diameter implies that such a C exists of length at most $\log n$.

The argument splits now according to whether π fixes some vertex in C . If it does not, then ϕ and ψ that map $[k]$ to C and to $\pi(C)$ respectively form a compatible type I configuration, and we are done. Otherwise, consider an arc $\Gamma = u \rightsquigarrow u'$ (possibly $u' = u$) of C so that: $v \in \Gamma$, and the only π -fixed points in Γ are u, u' . We obtain a compatible type II configuration by letting ϕ map $[k]$ to Γ and ψ map $[k]$ to $\pi(\Gamma)$. □

4 Connections with the Reconstruction Problem

The purpose of this section is to prove:

Theorem 13. *For every $0 \leq p \leq 1$ whp a $G(n, p)$ graph is reconstructible.*

We may clearly restrict ourselves to the range $0 \leq p \leq \frac{1}{2}$, since a graph is reconstructible iff its complement is reconstructible. We may further restrict our attention to the range $\frac{(1-\epsilon)\log n}{n} \leq p \leq \frac{(5/2+\epsilon)\log n}{n}$ since the theorem is known for the two complementary ranges. For $p \geq \frac{(5/2+\epsilon)\log n}{n}$ this was done by Bollobás [2]. Also, disconnected graphs are reconstructible [3], which takes care of the range $p \leq \frac{(1-\epsilon)\log n}{n}$. One further simplification is that for p in the above range, G almost surely has no $K_{3,2}$ subgraph. So we can and will be assuming this below. Our line of argument resembles the first part of the proof of Theorem 1. However, we need to adapt Lemma 8, a key step in that proof. This lemma gives an upper bound on $\Pr(\nabla(\phi(i)) = \nabla(\psi(i)))$, while here this equality gets replaced by an *approximate* equality as we now define.

For two multisets of integers we say that $A \approx B$ if they can be made equal by applying some of the following operations to each of them. (Here X refers to either A or B).

- Decrease some elements of X by 1 or 2. The total subtracted sum must be ≤ 4 .
- Delete one or two elements of X .

Definition 14. A configuration (ϕ, ψ) is acceptable if there exist vertex sets $U, W \subseteq V$ of size $n-2$ such that $\text{im}(\phi) \subseteq U$, $\text{im}(\psi) \subseteq W$, and G_U and G_W are isomorphic through a graph isomorphism π that maps $\phi(i)$ to $\psi(i)$ for every i .

Lemma 15. Whp, G contains no acceptable configurations of type I or II.

Proof. We first claim that $\nabla_G(u) \approx \nabla_G(\pi(u))$, for every $u \in U$ for U, W and π as above. This is so, since the property of π implies $\nabla_{G_U}(u) = \nabla_{G_W}(\pi(u))$. These are subgraphs of $n-2$ vertices and the effect of the two missing vertices is limited due to $K_{2,3}$ -freeness. Since G is $K_{3,2}$ free, $|N(u) \cap N(v)| \leq 2$ for every $v \in V \setminus U$. Hence, by removing v from G the possible changes in $\nabla(u)$ are: (i) Decreasing one or two elements of $\nabla(u)$ by 1: Each vertex in $N(u) \cap N(v)$ (of which there are at most two) may lose one neighbor, (ii) Removal of a single element from $\nabla(u)$ (the element corresponding to v itself, if $uv \in E$).

To prove the Lemma, we first strengthen Lemma 8, and replace the condition $\nabla(\phi(i)) = \nabla(\psi(i))$ by $\nabla(\phi(i)) \approx \nabla(\psi(i))$. The proof is essentially the same, with one change: Clearly the multiset $\{d(x, V \setminus \tilde{N}(T)) \mid x \in U_i\}$ is uniquely determined by the condition $\nabla(\phi(i)) = \nabla(\psi(i))$. Now we operate under the weaker condition $\nabla(\phi(i)) \approx \nabla(\psi(i))$. Rather than the above multiset, we consider a multiset where at most two of the entries are " $*$ " which stand for the possibly deleted vertices. This multiset can take on only $\text{poly}(np)$ possible values. Lemma 4 and a union bound argument yield:

$$\Pr(\nabla(\phi(i)) \approx \nabla(\psi(i))) \leq (np)^{-\Omega(\sqrt{np})} \cdot (np)^{O(1)} = (np)^{-\Omega(\sqrt{np})}.$$

By Lemma 7 and the stronger version of Lemma 8 proved here, the expected number of acceptable type I or type II configurations in G is at most

$$\sum_{k=3} \log n (np)^{-\omega_n(k)} \leq (np)^{-\omega_n(1)}.$$

□

Definition 16. We say that a vertex pair $u, v \in R(G)$ is interior if $R(G \setminus \{u, v\}) = R(G) \setminus \{u, v\}$.

Lemma 17. Whp, for every interior vertex pair $\{u, v\}$ it holds that (i) every automorphism of $G_{V \setminus \{x, y\}}$ fixes $R(G) \setminus \{u, v\}$ and (ii) For every interior vertex pair $\{x, y\} \neq \{u, v\}$, the graphs $G_{V \setminus \{u, v\}}$ and $G_{V \setminus \{x, y\}}$ are non-isomorphic.

Proof. We may assume that $\text{diam}(G) < \frac{\log n}{8}$, as this holds whp [4]. Also, $\text{diam}(G_{V \setminus \{u, v\}}) < \frac{\log n}{2}$ (likewise for $\{x, y\}$) since the removal of a vertex at most doubles the diameter. By Lemma 15, we may also assume that G has no acceptable type I or II configurations.

We prove both parts of the Lemma together by considering as well the case $\{x, y\} = \{u, v\}$. Assume that there exists an isomorphism π between $G_{V \setminus \{u, v\}}$ and $G_{V \setminus \{x, y\}}$ that does not pointwise fix the 2-core. To prove the Lemma, it is enough to show that there exists an acceptable type I or II configuration in G . We consider two cases, first where π moves some vertex in $R(G_{V \setminus \{u, v\}})$ that resides in a cycle. Since $\text{diam}(G_{V \setminus \{u, v\}}) < \frac{\log n}{2}$, we may assume that this cycle has length $< \log n$. The existence of an acceptable type I or II configuration follows from the same argument as that in Theorem 1. In the second case, π fixes pointwise every cycle of $R(G_{V \setminus \{u, v\}})$. Thus, it must map some path between two vertices in cycles, fixed by π , to a different path between these two vertices. Due to the bound on the diameter of $G_{V \setminus \{u, v\}}$, the length of these paths must be $< \frac{\log n}{2}$, which yields a type II acceptable configuration. \square

Proof of Theorem 13. We may and will assume that G satisfies the conclusion of Lemma 17. For $u \in U$, we denote $\tilde{G}_u := G_{V \setminus \{u\}}$.

We claim that the cardinality $|R(G)|$ is reconstructible. Indeed, it is known [3] that the degree sequence of G is reconstructible, and thus, the property $R(G) = V$ is recognizable. Now, assume that $R(G) \neq V$. It is clearly possible to determine $d(u)$ from \tilde{G}_u . Also $R(G) = R(\tilde{G}_u)$ when $d(u) = 1$. Since $u \in R(G)$ iff $|R(\tilde{G}_u)| < |R(G)|$, we can determine whether $u \in R(G)$ by observing \tilde{G}_u .

We also note that the degree sequence of G 's 2-core is reconstructible. Indeed, if $V = R(G)$ this follows from the reconstructibility of G 's degree sequence. Otherwise, the 2-core itself is reconstructible, as above.

Let $A = \{u \in R(G) \mid d(v, R(G)) \geq 4 \text{ for all } v \in N(u)\}$. Note that every vertex pair in A is interior. It is not hard to determine whether $u \in A$ given \tilde{G}_u , based on the reconstructibility of the 2-core's degree sequence. We claim that A contains almost all vertices. By Lemma 5, there holds whp $|R(G)| \geq n - o(n)$. For $v \in V$

$$\Pr(d(v, R(G)) \leq 3) \leq O(np^3 e^{-np})$$

and by the union bound

$$\Pr(\exists v \ v \in N(u) \wedge d(v, R(G)) \leq 3) \leq O(np^4 e^{-np}) \leq o(1).$$

So, let $v' \in V \setminus \{u\}$ and $u' \in V \setminus \{v\}$ be such that $\{u, v'\}$ and $\{v, u'\}$ are interior pairs, and there exists an isomorphism π between $G_{V \setminus \{u, v'\}}$ and $G_{V \setminus \{u', v\}}$. By Lemma 17, this holds only when $u = u'$, $v = v'$ and π fixes the 2-core pointwise. Using this property, we can identify the vertices v and u respectively in the graphs \tilde{G}_u and \tilde{G}_v and identify each vertex in the 2-core of one graph with its counterpart in the other. This allows us to reconstruct G up to the question of whether $uv \in E$. Since $|E|$ is reconstructible, this last question can be answered as well. \square

5 Connections with the Canonical Graph Labeling Problem

In this section we describe a polynomial time random graph canonical labeling algorithm for graphs in $G(n, p)$ where $\omega(\frac{1}{n}) \leq p \leq n^{-(0.5+\epsilon)}$.

Let \mathcal{C} be the collection of all *rooted, oriented* cycles of length $3 \leq k \leq \log n$ in an n -vertex graph $G = (V, E)$. We use $A \leq B$ to denote the lexicographic ordering between multisets of integers, where the elements in A and in B appear in increasing order. We equip \mathcal{C} with the semi-order \prec where short cycles precede longer ones. For two cycles $X = (x_1, \dots, x_k), Y = (y_1, \dots, y_k) \in \mathcal{C}$ we say that $X \prec Y$ if for some i there holds $\nabla(x_i) < \nabla(y_i)$ and $\nabla(x_j) = \nabla(y_j)$ for every $1 \leq j < i$.

We claim that for the relevant range of p , a $G(n, p)$ graph satisfies the following conditions whp:

1. Each connected component of $G_{V \setminus R(G)}$ is a tree of size $\leq \log n$.
2. $\text{diam}(G) < \frac{\log n}{2}$.
3. $\mathcal{C}(G)$ is totally ordered by \prec .

Property (1) is easy to derive by a first-moment argument. For property (2), see [4]. A proof of Property (3) follows from property (2) by a simple variation of the proof of Theorem 1.

We now explain how to canonically label a graph $G = (V, E)$ with these three properties. To a vertex v that is contained in a cycle we assign the \prec -smallest label $X = (x_1, \dots, x_k) \in \mathcal{C}$ over all cycles for which $v = x_1$. This label can be found in polynomial time. If k is the length of the shortest cycle through $v = x_1$, then it is easy to show that there are at most n^3 such cycles. We scan all of them and pick the \prec -smallest one.

Note next, that a vertex $v \in R(G)$ that is not contained in a cycle must reside on the unique path between two vertices $u, w \in V$, each contained in a cycle. Therefore v is uniquely defined by its distances from u and from w . This, and the labels of u and w , give us a unique label for v .

Finally we find labels for vertices in $V \setminus (R(G))$. By property (1), such a vertex belongs either to (i) a tree of size $\leq \log n$ rooted at some vertex of $R(G)$ or (ii) an acyclic connected component of size $\leq \log n$. Let v be a vertex of type (i), belonging to a tree T rooted at $u \in R(G)$. There are only $\text{poly}(n)$ rooted trees of size $\leq \log n$ [8], so we can list them and give a unique polynomial-length label to each vertex of each such class. We label v by a pair (x, y) , where x is the label of the vertex corresponding to v in T 's isomorphism class in the list, and y

is u 's label. That is, v is labeled as "The vertex of type x in the tree rooted at u ". Type (ii) vertices are likewise handled, using a list of all isomorphism classes of non-rooted trees. To deal with vertices on acyclic connected components, collect all connected components of the same isomorphism class and give each of them a unique number. The label of v consists of the type of tree that contains it, that tree's ordinal number in its isomorphism class, and v 's location in that tree.

6 Discussion and Open Problems

For smaller values of p the structure of $\text{aut}(G)$ may become somewhat more complicated. For $p = \Theta(\frac{1}{n})$, a $G(n, p)$ graph has, with probability bounded away from zero and one, some small symmetric components, e.g., an isolated triangle. Moreover, with probability $\in (0, 1)$ even the 2-core of the graph's giant component, has a nontrivial symmetry. This may result e.g., from a triangle that "hangs off" the 2-core. However, as shown in [9], whp this 2-core has a unique biconnected component of $\Omega(n)$ vertices. We suspect that this giant biconnected component is rigid whp.

For $\frac{(5/2+\epsilon)\log n}{n} \leq p \leq \frac{1}{2}$ it was shown by Bollobás [2] that not only is G reconstructible whp, such graphs have reconstruction number three. We do not know whether this holds as well for smaller and substantially smaller values of p .

References

- [1] L. Babai, P. Erdős, and S. M. Selkow, *Random graph isomorphism*, SIAM Journal on Computing **9** (1980), no. 3, 628–635.
- [2] B. Bollobás, *Almost every graph has reconstruction number three*, Journal of Graph Theory **14** (1990), no. 1, 1–4.
- [3] J. A. Bondy, *A graph reconstructor's manual*, Surveys in combinatorics, 1991, 1991.
- [4] F. Chung and L. Lu, *The diameter of sparse random graphs*, Advances in Applied Mathematics **26** (2001), no. 4, 257–279.
- [5] T. Czajka and G. Pandurangan, *Improved random graph isomorphism*, Journal of Discrete Algorithms **6** (2008), no. 1, 85–92.
- [6] F. Harary, *On the reconstruction of a graph from a collection of subgraphs*, Theory of graphs and its applications (proc. sympos. smolenice, 1963), 1964, pp. 47–52.
- [7] P. J. Kelly, *A congruence theorem for trees*, Pacific Journal of Mathematics **7** (1957), no. 1, 961–968.
- [8] R. Otter, *The number of trees*, Annals of Mathematics (1948), 583–599.
- [9] B. Pittel, *On tree census and the giant component in sparse random graphs*, Random Structures & Algorithms **1** (1990), no. 3, 311–342.

- [10] S. M. Ulam, *A collection of mathematical problems*, Interscience, 1960.
- [11] E. M. Wright, *Graphs on unlabelled nodes with a given number of edges*, Acta Mathematica **126** (1971), no. 1, 1–9.